Citizen Scientists Are Not Just Quiz Takers: Information about Project Type Influences Data Disclosure in Online Psychological Surveys

ANNA RUDNICKA [iD]

SANDY GOULD [iD]

ANNA COX [iD]

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Traditionally, citizen science has centred on giving lay people opportunities to learn about science by participating in it. Lately, psychological citizen science projects have increasingly aimed to attract participants by providing an opportunity for self-learning. Unfortunately, these citizen science projects sometimes resemble quizzes hosted by private companies that have been criticised for suboptimal privacy practices. We were interested in whether similarity to these dubious quizzes could jeopardise the ability of citizen science projects to collect sensitive personal data. We found that people who join an online quiz disclose a lower volume of sensitive data than those who join a citizen science project. Moreover, post-recruitment encouragement to learn about science influenced data disclosure among quiz takers but not among citizen scientists. These findings have implications not only for the effective design of citizen science but also for the broader debate about consent to disclose data in research, academic or otherwise.

CORRESPONDING AUTHOR:

Anna Rudnicka

University College London, GB

anna.rudnicka.15@ucl.ac.uk

INTRODUCTION

Online quizzes have recently attracted negative attention. For example, a company called Cambridge Analytica is reported to have gathered data through an online personality quiz "thisismydigitallife," allegedly aiding the delivery of targeted political advertisements to both Facebook users and also to their friends who never took the quiz themselves (Hern 2018). This excessive access to personal data was later described as a "breach of trust" by the CEO of Facebook, Mark Zuckerberg (Zuckerberg 2018), demonstrating how online personality quizzes can have consequences that reach far beyond allowing people to learn something about themselves. Writing about this episode, Paul Bisceglio noted that "online personality analysis can easily blur the line between opting in and opting out" (Bisceglio 2017).

Appropriate consent and a sense of fairness are important not only from the standpoint of ethics, but also to support data collection. Studies have demonstrated that, in consumer contexts, people withhold information when they perceive the reward for disclosure to be less than the costs of disclosure and therefore feel that the disclosure situation that they are in is unfair (Malheiros, Preibusch, and Sasse 2013; Sheehan and Hoy 1999).

Citizen scientists contribute their time and effort without remuneration. However, participation in citizen science can offer different non-monetary benefits. These include altruism, learning, contributing to a larger effort, and being part of a group of people working towards the same goal; recruitment messages that emphasised these four benefits have been shown to have different impacts on people's willingness to participate in citizen science (Lee et al. 2018). Studying citizen scientists' willingness to disclose personal data, Rudnicka, Cox and Gould (2019) adapted these four messages for the context of a psychological survey. They found that people who saw a message encouraging them to learn about science, at the start of the survey, disclosed a higher volume of data than those who saw other messages.

These findings suggest that the way citizen scientists are encouraged to take part in a project may influence both whether they join or contribute, as well as their willingness to share personal data. In cases of projects that rely on the collection of personal data, it could be advisable for coordinators to identify what motivates their target population to share data, to design the project such that it delivers these benefits, and to use appropriate messaging when advertising the project. This could support a sense of fairness among the participants, while also directing the focus of coordinators to what they give back to their participants.

A sense of fairness is particularly important when citizen scientists are asked to share personal data. Although traditionally citizen science has been focused on the contribution that lay persons can provide to research through collection or analysis of data, the burgeoning field of psychological citizen science encourages participants to contribute data about themselves. One online platform, LabintheWild hosts multiple quizzes that allow citizen scientists to discover a personal score while contributing to scientific research. While this benefits citizen scientists by increasing their self-knowledge, this format is uncomfortably similar to online quizzes hosted by private companies that seek to covertly extract data from the participants.

As an increasing number of citizen science projects, especially in psychology and human-computer interaction, resemble online quizzes hosted by private companies, this poses an ethically salient question: Are people able to distinguish between legitimate citizen science projects hosted by researchers, and online quizzes created by companies for surveillance purposes? Furthermore, do individuals disclose personal data more readily when they know that they are taking part in citizen science, rather than in a quiz hosted by a private company—even when in both cases they are prompted to discover something about themselves?

In this study, we compared two disclosure environments: one in which an online survey was advertised as a "discover your score" quiz, and one in which the same survey was advertised as a citizen science project. We aimed to shed light on whether recruitment-stage framing of a project can influence the volume of disclosed sensitive data, as it had been shown to impact participant recruitment and volume of contribution to a project (August et al. 2018). We also aimed to clarify whether the way in which an online survey is advertised could mediate the previously reported (Rudnicka et al. 2019) relationship between post-recruitment messages about the benefit of participation, and data disclosure among citizen scientists.

The results of our study offer insights into how information, communicated to participants at different stages of an online psychological survey, can impact willingness to share data. In particular, we demonstrate that recruitment-stage information about project type directly influences the volume of disclosed sensitive data as well as mediating the relationship between post-recruitment messages about the benefit of participation and data disclosure. Our findings can help ensure that citizen science research studies are able to collect sustainable volumes of personal data at a time of increased public scrutiny in the areas of privacy and data protection.

DATA COLLECTION IN CITIZEN SCIENCE: LEGISLATIVE CHANGES

Citizen science involves collaboration between researchers and laypersons who voluntarily contribute to research activities through, for example, gathering or analysing data (Silvertown 2009). Participants are sometimes asked to disclose personal data or to provide information that, while not focused on citizen scientists as data subjects, may still include sensitive metadata such as their location, time of data collection, or IP address. Research has shown that, in the case of experienced, long-time participants, the same reasons that prompt people to take part in citizen science—for example a sense of community—can also encourage them to share sensitive data, even when they have some concerns about data privacy (Bowser et al. 2017).

Recent years have seen a rise in public concern with the privacy and security of data transmitted over the Internet (Ofcom 2017). In parallel, the new EU-wide privacy laws that became binding in all member states in May 2018 reflect a new regulatory approach to protecting the privacy of individuals, one that emphasises informed consent and clear communication (European Parliament and Council of European Union 2016). Stricter privacy laws necessitate increased clarity in communication with individuals when encouraging them to disclose personal data. To ensure that citizen science projects are ethically sound, participants ought to be encouraged towards disclosure not through dark patterns (Bösch et al. 2016) but rather through honest and transparent communication that realistically describes the benefits of taking part in citizen science. While citizen scientists may decide to share their commodity—personal information—because of the potential gains that participation in a citizen science project could bring them—be it learning, satisfaction that they are supporting scientific progress, or a sense of belonging that stems from being part of a social group with a common goal (Lee et al. 2018)—the increased scrutiny, legal and societal, of the ways institutions gather and use sensitive data requires citizen science coordinators to place more emphasis on informed consent. This, in turn, could make it harder to recruit citizen scientists who are willing to share their data, because if projects fail to offer and effectively communicate the attractive benefits of participation, participants may not view data disclosure requests as fair.

COLLECTION OF DATA ABOUT THE CITIZEN SCIENTIST

Many citizen science initiatives involve the sharing of personal data; for example, conservation projects ask people to provide ecological observations accompanied by time, date, and location, which could help infer their whereabouts, walking patterns, or even, potentially, home address. However, projects that recruit citizen scientists to participate in psychological research, an area referred to as citizen psych-science (Jennett et al. 2014), are reliant on people's willingness to overtly share data about themselves, rather than things in the world around them. Examples of such projects include "Mappiness" (*https://www.mappinessapp.com*), which asks volunteers to record their mood in different locations to discover how different environments affect people's happiness; "Errordiary" (*www.errordiary.com*), for which participants report their everyday errors on the social platform Twitter to help scientists understand the types of errors that people deal with in their lives: and the "Emotional Brain Study" (*https://sites.google.com/view/emotionalbrain/home*), which encourages citizen scientists to complete memory and attention tasks, and to track their mood, in support of research into cognitive tracking interventions. Citizen scientists could become increasingly important survey participants, as researchers using the survey methodology report that paid studies are vulnerable to bots (Perkel 2020). While one way to avoid non-genuine participation that can be rife in paid studies could be restricting recruitment to university recruitment pools, this could in turn negatively influence the diversity of participant samples.

Bowser et al. (2014) highlight the importance of protecting citizen scientists' privacy and argue that projects should be designed not only with legal and regulatory implications in mind, but also considering ethical best practice. One of their recommendations for project hosts is to collect only minimum personal data about the volunteers themselves. While this is an important consideration (applicable, for example, to conservation and human computation citizen science projects), it may be hard to implement in the case of citizen psych-science in which the citizen scientists are also the focus of research.

Psychological citizen science projects will often necessarily involve requests for information legally defined as special category personal data (for example, health history). Even data that falls outside of the bounds of that classification, such as questions about people's money or friends, can be considered sensitive or private. The first experimental study of data disclosure in citizen science has demonstrated that individuals are significantly more likely to share information when asked about a neutral topic (in that case, their sleep habits and preferences) than sensitive or personal topics such as family history, bill payments, or friends on social media (Rudnicka et al. 2019). While we acknowledge that researchers should avoid collecting personal data unless it is necessary for the success of

the project, and that close ethical examination of risks and benefits as well as rigorous data protection practices should always be prioritised, psychological research is likely to include at least some queries that could be perceived as personal or sensitive.

Because psychological research requires data about the participants themselves, many citizen psych-science studies rely on one-off contributions from a large volume of participants, which introduces another complication. Individuals giving the same information about themselves over and over would obviously not be helpful, whereas a single participant can classify, say, thousands of galaxies or record hundreds of bird observations. As noted by Wiseman et al. (2014), the short window of participation in one-off citizen science surveys precludes the development of a sense of community, which is known to facilitate data disclosure in long-term projects (Bowser et al. 2017).

Nevertheless, research suggests that the way we encourage people to take part in projects can influence their willingness to share data. In a recent study of citizen psych-science, presentation of short messages emphasising different benefits of participation influenced the volume of disclosed data; participants who saw a message encouraging them to learn about science disclosed a larger volume of data than participants encouraged through other messages (Rudnicka et al. 2019).

CITIZEN SCIENCE SURVEYS VERSUS ONLINE QUIZ SURVEYS

Both the popular press and academic publications have highlighted the ethical concerns associated with online quizzes. For example, personality quizzes can lead to the disclosure of data that may not appear sensitive but can render a person vulnerable to manipulation, as was the case in the Cambridge Analytica scandal (Susser, Roessler, and Nissenbaum 2019). Another example of the type of online quizzes that extract a lot of information in a way that may not be clear to all users is the immensely popular quiz format used by the Buzzfeed (*https://www.buzzfeed.com*) website. Some Buzzfeed quizzes can appear innocuous, such as "Customize Your Cheesecake And We'll Reveal Which High School Clique You Belong In" (Aspler 2019), while others include very invasive questions (the "How Privileged Are You?" quiz asks people to check off all the statements that apply to them, including "I have never been raped" and "I have never attempted suicide" (Thomas and Jha 2017). Such data, sometimes supplemented by other information like the person's gender or email address, are collected by the website

for research and marketing purposes, and while the data are said to be anonymised and aggregated, some commentators still consider them a cybersecurity risk (Barker 2014). The privacy statements of many websites that provide such learn-something-about-yourself quizzes reserve the right to aggregate data and share them with third parties (Harding 2018).

The market-driven approach that prompts private companies to host quizzes online in an attempt to extract as much data as possible from the people who fill them in, is different from the atmosphere of co-operation between citizen scientists and project hosts (Bowser et al. 2017) that facilitates the sharing of personal data in collaborative research projects. Nevertheless, academic researchers increasingly employ the quiz format to encourage volunteer participation in research studies, emphasising opportunities to test yourself or to learn something about yourself (often by discovering a personal score). Benefits to participants, like self-discovery, are an important part of any ethical calculus, but framing citizen science projects as solely or mostly an opportunity for self-discovery does not give participants an opportunity to strive towards a common goal, and therefore does not encourage collaborative data sharing. It is therefore important to examine whether the context—citizen science or a quiz to learn about oneself—in which participants are recruited for an unpaid online survey may influence these participants' willingness to share personal data.

LEARNING ABOUT SCIENCE OR LEARNING ABOUT YOURSELF?

Citizen science projects play an increasingly important role in scientific research, not only as a means of collecting or analysing research data, but also as an educational tool (Bonney et al. 2009), with the ability to learn about the scientific process or an area of interest considered an important benefit of participation (Theobald et al. 2015; Domroese and Johnson 2017; Oberhauser and LeBuhn 2012).

Self-discovery, however, has been associated with the "quantified self" movement, in which laypersons gather data about themselves in the quest for self-knowledge. Wiseman et al. (2017) define citizen science as distinct from the quantified self movement, where "[u]nlike CitSci, the motivation for collection of data comes not from the "selfless" act of participating in an experiment belonging to someone else, but rather from the opportunity to learn more about oneself" (p.2).

Nevertheless, some citizen science projects motivate participation through a self-discovery framework. The LabintheWild (*www.labinthewild.org*) platform, for example,

hosts studies in the area of cognition ("How well can you find patterns?" "Do you make assumptions about people without knowing it?"), psychology ("What is your personality?") or human-computer interaction ("Amazon, Apple, Facebook, Google: Can you tell the difference?").

Research has shown that advertising the opportunity to self-learn can boost recruitment into unpaid online experiments (August et al. 2018). This suggests that citizen science coordinators might be right to use this framing to advertise projects. Nevertheless, success in recruitment does not always correlate with success in data collection. In the study conducted by Rudnicka at al. (2019), a message that emphasised learning about science was most effective, out of four different messages, at eliciting data disclosure, but second-least effective at recruitment (Rudnicka et al. 2019). The impact on data disclosure of prompting participants to learn about themselves rather than about science has not yet been assessed.

CURRENT STUDY

To the best of our knowledge, this study was the first experimental comparison of online data disclosure between citizen scientists and quiz takers. We set out to compare data disclosure behaviour in two settings: an online citizen science survey, and an online quiz that provides participants with a personal score. In both contexts, we explored the impact of post-recruitment messages emphasising either learning about science (more aligned with citizen science projects) or learning about oneself (more aligned with online quizzes). This allowed us to compare the relative impact on data disclosure of information presented at the point of recruitment, and information presented at the start of the survey. That is, we gained insight into the relative importance of recruiting a particular sample of participants versus presenting adequate information to already recruited participants, insofar as it affects the ability of citizen science projects to collect personal data.

We set out to test two hypotheses: H1: Participants who fill in a survey advertised as citizen science will disclose a larger volume of sensitive data than participants who fill in a survey advertised as an online quiz. H2: The relationship between the post-recruitment messages and the disclosure of sensitive data will be moderated by project type.

METHOD
PARTICIPANTS

Participants were recruited online via Twitter, in two rounds, and participation was open to all individuals over the age of 18, irrespective of location. First, we recruited for the citizen

science group, tweeting: "Take part in our #CitizenScience survey 'Sleep Patterns'!" Through paid advertisements we popularised this tweet among people using relevant hashtags such as #citizenscience or #psychology. The tweet was accompanied by a photo of a sleeping cat. Out of 297 individuals who decided to take part in the Sleep Patterns survey in that recruitment round, 142 dropped out during the survey. Only data provided by participants who reached and completed the last survey question (n = 155) were analysed. Although respondents were randomly assigned to three message groups in equal numbers, due to the pattern of participant attrition, the final sample included uneven numbers of participants in these three groups: "Learn about science'" (n = 57), "Learn about self" (n = 62), and "Control" (n = 36).

We then recruited participants for the online quiz group, tweeting: "Discover your Sleep Score in our #quiz 'Sleep Patterns'!" We popularised this tweet through paid advertisements among people using relevant hashtags such as #quiz or #psychology. The tweet was accompanied by the same photo of a sleeping cat. We held the recruitment open until the same number of participants (n = 155) had reached and completed the final question of the survey. The final sample of 155 participants originated from 450 individuals who started the survey, with 295 dropping out. Although respondents were randomly assigned to three message groups in equal numbers, due to the pattern of participant attrition, the final sample included uneven numbers of participants in these three groups: "Learn about science" (n = 56), "Learn about self" (n = 51), and "Control" (n = 48).

The total sample consisted of 310 participants. All but two reported their age, with ages ranging from 18 to 75 years old (mean = 40, SD = 15). Data about gender, as reported by the participants (available for n = 310) were as follows: n = 177 female, n = 122 male, n = 7 non-binary, n = 1 preferred not to answer, and n = 3 answered in their own words. All participants reported Internet usage: n = 152 reported using the Internet "all the time," n = 148 reported using the Internet "several times per day," n = 8 reported using the Internet "most days," and n = 2 reported using the Internet several times per week." Only a small proportion of participants reported having had previous experience with citizen science (n = 17, "once;" and n = 12, "more than once"), while the majority reported either not having any citizen science experience (n = 225) or being "not sure" (n = 55); data for one participant were missing.

MATERIALS

The survey was hosted on the Qualtrics (*https://www.qualtrics.com*) platform. To ensure that we were able to investigate

spontaneous and authentic disclosure behaviour, the study involved deception. The Participant Information Sheet stated that the aim of the research was "to learn about how different sources of stress in a person's life are related to their sleep patterns." Participants were explicitly informed that research designs often require that the full intent of the study not be explained prior to participation. After filling in the survey, the participants were provided with a debriefing message which explained that the aim of the research had been to study data disclosure behaviour. Participants were also able to withdraw their data from the study.

The survey consisted of the following elements: a post-recruitment message (1 out of 3, randomly assigned); a Participant Information Sheet, comprising seventeen consent questions compliant with the General Data Protection Regulation (European Parliament and Council of European Union 2016) and the Data Protection Act (UK Government 2018); demographics questions (age, gender, Internet use, previous citizen science participation); 19 neutral items, and 14 sensitive items. The post-recruitment messages were as follows: 1) "Extend your knowledge of health psychology by participating in the Sleep Patterns survey!" (learning about science), 2) "Fill the Sleep Patterns survey and discover your sleep score!" (learning about self), and 3) "Welcome to the Sleep Patterns survey!" (control group).

The Morningness-Eveningness Questionnaire (Horne and Östberg 1976) served as Neutral Items; the response mode was modified to suit the online format of the study, and participants were able to obtain a personal sleep score immediately after completing the survey. The Sensitive Items had originally been drawn from a list of Novel Items in a study of data disclosure among credit card applicants (Malheiros et al. 2013). To validate the use of these questions, we conducted a pilot card sorting study (n = 12).

PILOT CARD SORTING STUDY TO VALIDATE MATERIALS

We aimed to test the assumption that the Sensitive Item questions would be interpreted as requesting sensitive data, and that the Neutral Item questions would be interpreted as requesting neutral data. We recruited 12 participants from among the graduate students in the department. They completed the task individually, on a researcher's laptop computer. The card-sorting task was hosted on the Optimal Workshop platform and began with instructions on the screen: "Citizen scientists are people who volunteer their time for research projects. One type of citizen science project is an online survey. Some surveys include questions that request sensitive data. Imagine that you are filling

a citizen science survey that looks at the connection between sleep habits and stress. You will see a list of 33 questions and 4 categories (Definitely Neutral, Somewhat Neutral, Somewhat Sensitive, Definitely Sensitive). Please assign the questions to categories. There are no right or wrong answers—we are interested in your opinions. For example, if you read a question, and think "this question asks me to share neutral data," then please assign it to one of the neutral categories. If you read a question and think "this question asks me to share sensitive data," then please assign it to one of the sensitive categories." Following instructions, participants clicked an arrow and saw two lists of items. On the left, they saw a list of cards with Neutral and Sensitive Items presented in random order. On the right, they saw a list of 4 categories, which represented 4 levels of sensitivity: Definitely Neutral, Somewhat Neutral, Somewhat Sensitive, and Definitely Sensitive. Participants were then shown a brief instruction message: "Take a look at the list of items on the left. Please sort those items into the groups provided on the right." They then sorted the questions until all questions were assigned to a category. After completing the task, participants were verbally debriefed.

We found that all of the Neutral Item questions were identified by the majority of participants as either Somewhat Neutral or Definitely Neutral. Similarly, all of the Sensitive Item questions were identified by the majority of participants as either Somewhat Sensitive or Definitely Sensitive. These findings confirm that these questions are appropriate for the study of disclosure decisions made in response to neutral and sensitive queries. The results of the pilot study allowed us to specifically examine the disclosure of sensitive information by focusing data analysis on participants' responses to the Sensitive Item questions.

DESIGN

The experiment used a 2x3 between-subjects design. The independent variables were Project Type and Message. Project Type had two levels (Citizen science, Online quiz) and was operationalised by recruiting participants with two different Twitter messages, one referring to citizen science, and the other referring to an online quiz. Message had three levels (Learning about science, Learning about self, Control). It was operationalised by presenting the participants with one of the three, randomly assigned, messages at the start of the survey. The dependent variables were Neutral Data Disclosure and Sensitive Data Disclosure, operationalised by asking people to answer, respectively, 19 Neutral Item questions and 14 Sensitive Item questions. In this study, we specifically focused on clarifying the impact of independent variables on Sensitive Data Disclosure.

PROCEDURE

Following the link advertised on the social network Twitter, the participants were redirected to the online survey hosted on the Qualtrics platform. Before commencing the survey, they were presented with one of the three, randomly assigned, messages. They were then asked to read the Participant Information Sheet and decide whether they wanted to take part in the study and fill in the consent from. Following consent, the participants filled in the survey, comprising 4 Demographics questions, 19 Neutral Item questions, and 14 Sensitive Item questions. After completing the survey, all participants were shown the debriefing message and they were presented with their personal sleep score, calculated based on their responses to the Neutral Items drawn from the Eveningness-Morningness Questionnaire (Horne and Östberg 1976), alongside online resources on staying safe online and managing sleep. The debriefing message was additionally emailed to the participants following the completion of the survey.

RESULTS

DATA DISCLOSURE ACROSS NEUTRAL AND SENSITIVE ITEMS

All participants in this study provided a response to every one of the 19 Neutral Item questions, resulting in 100% neutral data disclosure in this study. Disclosure of sensitive data in this study ranged from 0 to 13 disclosed items (out of 14), with a mean of 4.93 items (SD = 1.96), equivalent to 35.2% sensitive data disclosure. As the counts were uneven in this study (19 Neutral Items and 14 Sensitive Items), they were transformed into percentages before a statistical comparison between the two was conducted. The result of a paired t-test (t = 81.644, df = 309, p = .000, 2-tailed) suggested that the proportion of disclosed Neutral Items was significantly larger than the proportion of disclosed Sensitive Items.

IMPACT OF PROJECT TYPE ON SENSITIVE DATA DISCLOSURE

We examined the impact of how the online survey was advertised to the participants (irrespective of the post-recruitment messages they later saw) in the two project type groups. In the citizen science group (n = 155), participants disclosed a mean of 5.2 Sensitive Items (SD = 2.24). In the online quiz group (n = 155), participants disclosed a mean of 4.66 Sensitive Items (SD = 1.59). A one-way between-subjects Analysis of Variance showed that the impact of project type on sensitive data disclosure was significant (F(1,308) = 6.045, p = .014, η = .019),

suggesting that individuals recruited for a citizen science project tend to disclose a larger volume of sensitive data than individuals recruited for an online quiz.

IMPACT OF MESSAGE ON SENSITIVE DATA DISCLOSURE ACROSS PROJECT TYPE GROUPS

This study aimed to clarify whether the relationship between messages that emphasise the benefit of participation and the disclosure of sensitive data varies between two types of projects: one advertised as a citizen science project, and a one advertised as an online quiz. Across the full sample of participants (n = 310), differences in disclosure between Message groups were not significant (F(2,307) = 2.099, p = .124, = .013).

There was a significant interaction effect between the variables of Message and Project Type (F(2,304) = 3.568, p = .029, η = .023), suggesting that post-recruitment messages may differently impact people's willingness to share sensitive data, depending on which project these participants were initially recruited for. We therefore proceeded to analyse the data from two project types separately, as we were interested in examining participants' data disclosure behaviour in these two distinct scenarios. *Table 1* and *Figure 1* showcase the mean disclosed Sensitive Items across the two Project Type groups: Citizen Science and Online Quiz.

In the Citizen Science group (n = 155), the mean disclosure of sensitive data across the message groups was lowest for the "Control" group, followed by "Learn about science," and was highest in "Learn about self." However, a One-Way Independent Analysis of Variance was not significant (F(2,152) = 2.276, p = .106, η = .029), suggesting that post-recruitment messages did not appear to impact participants' disclosure behaviour if they were initially recruited for a citizen science project.

In the Online Quiz group (n = 155), the mean disclosure of sensitive data across the Message groups was highest in "Learn about science," followed by "Control" and "Learn about self." A One-Way Independent Analysis of Variance (F(2,152) = 3.472, p = .034, η = .044) suggested that post-recruitment messages did appear to impact participants' disclosure behaviour if they were initially recruited for an online quiz. A post hoc Tukey test (p = .036) suggested that the difference in sensitive data disclosure between participants in the "Learn about science" group and participants in the "Learn about self" group was statistically significant. These results suggest that encouraging quiz takers to learn about science is more effective at eliciting the disclosure of sensitive data than encouraging them to learn about themselves.

MESSAGE				
		LEARN ABOUT SCIENCE	LEARN ABOUT SELF	CONTROL
Project Type				
Citizen Science	N	57	62	36
	M	**5.11**	**5.61**	**4.64**
	SD	1.99	2.68	1.59
Online Quiz	N	56	51	48
	M	**5.09**	**4.33**	**4.50**
	SD	1.93	1.44	1.17

Table 1 Mean disclosed Sensitive Items (*bold*) across Project Type groups.
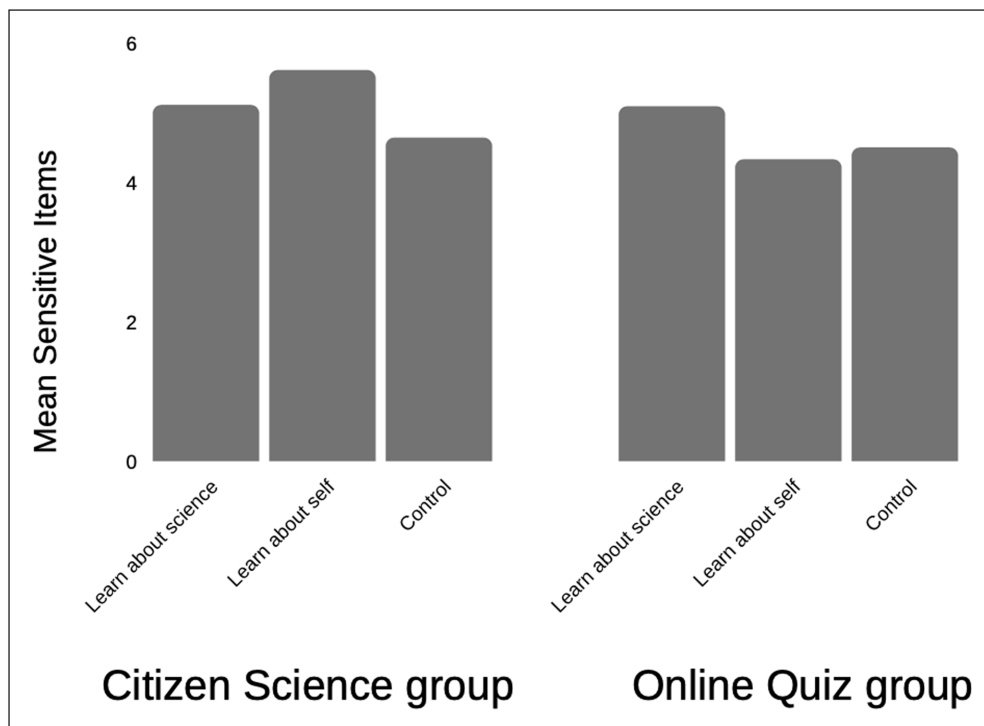
Figure 1 Mean disclosure of sensitive data across the three Message groups in the Citizen Science group and Online Quiz group.

DISCUSSION
DATA DISCLOSURE
All of the participants in this study provided an answer for every Neutral Item question, contrasted with mean disclosure of 4.93 (out of 14) items in the case of Sensitive Item questions. This demonstrates, in line with previous experimental research into data disclosure among citizen scientists (Rudnicka et al. 2019), that when faced with requests for neutral data in online surveys, people readily provide information. Moreover, it is clear that individuals taking part in this survey were able to distinguish between neutral and sensitive data requests. These findings highlight the need for awareness, on the part of citizen

science coordinators, of the sensitivity of the data requests included within citizen science projects.

CITIZEN SCIENTISTS VERSUS QUIZ TAKERS
This study involved two modes of recruitment: one that asked people to take part in an online quiz and find out their sleep score, and one that asked them to participate in a citizen science survey. Although both groups filled in the same questionnaire, which made a mention of citizen science (the last demographics question asks: "Have you ever taken part in a Citizen Science project before?"), people recruited for a citizen science survey disclosed a larger volume of sensitive data than the people recruited for

an online quiz. Moreover, while we ensured that the final participant numbers in both groups were even (n = 155 in each group), the participant attrition was greater in the online quiz group (a primary sample of n = 450 signed up for the study) than in the citizen science group (a primary sample of n = 297 signed up for the study).

These findings suggest that when people decide to fill in an online survey because they wish to take part in citizen science, they are more likely to share sensitive data within the project and less likely to drop out.

Two types of implications follow from these findings. Firstly, these findings have practical implications relevant for the design of online research projects and the study of citizen science. It appears that emphasising the opportunity for self-discovery at the recruitment stage is a less useful strategy to recruit participants into citizen science surveys that involve the disclosure of sensitive data, when compared with emphasising the brand of citizen science. It is possible that people associate citizen science projects with trustworthy public institutions and therefore feel more comfortable sharing information when the concept of citizen science is emphasised. In experimental studies, researchers often focus on the impact on antecedents of trust, such as the brand or appearance of the website, or the privacy policy, on the willingness to disclose data (Gómez-Barroso 2018). Experimental research into the impact of trust in data disclosure among citizen scientists should be conducted to determine the possible role of the citizen science brand as an antecedent of trust and its impact on data disclosure. Furthermore, it is also possible that individuals focused on self-discovery may prioritise their own goals and protect personal data when possible (privacy protection behaviours have been observed in people's interaction with private companies [Malheiros, Preibusch, and Sasse, 2013]). It is also plausible that individuals focused on self-discovery assess data requests more critically, honouring only the ones that appear relevant to their goal (discovery of a personal score). In contrast, participants prompted to learn about science may, when focused on learning and therefore the limits of their existing knowledge, be more open minded about which data requests should be included in the study. Secondly, these findings have ethical implications: It appears that mentioning the citizen science "brand" at the stage of recruitment could lead people to share a larger volume of sensitive data than they would if they were approached to fill in a quiz or survey that is not a legitimate citizen science project. The breadth and variety of citizen science initiatives (Auerbach et al. 2019) can make it hard to delineate specific boundaries for what citizen science is or is not. However, a lack of agreed-upon boundaries might make it possible for bad actors to obtain sensitive data

from people simply by advertising a quiz or other disclosure request as citizen science. This puts a responsibility on citizen science coordinators, whose institutions stand to benefit from the volunteer contributions of citizen scientists, to help ensure that the brand of citizen science is not misused. This could be done by developing a system of certification, or by helping the wider public understand how to check the authenticity of a project. There are existing guidelines outlining good standards for designing and conducting citizen science projects. For example, the European Citizen Science Association lists on their website ten principles developed by the "Sharing best practice and building capacity" working group, which include consideration of legal and ethical issues (European Citizen Science Association 2015). However, as regulators around the world change laws to accommodate growing societal concerns about data privacy, it will become important to work towards a uniform framework for ethical and effective data collection among citizen scientists to support the wide variety of citizen science projects.

It is important to note that while the citizen science mode of recruitment was more effective than the online quiz alternative, quiz takers also engaged in the disclosure of sensitive data. It is possible that this was because other parts of the survey, such as the Participant Information Sheet, noted the academic host institution and included a consent form, both features of legitimate academic studies. Nevertheless, these features could have been easily manufactured by bad actors causing participants to render themselves vulnerable. With the prevalence of online scams, as well as corporate surveillance, it is important to conduct further research into how and why quiz takers disclose personal data online. This is especially true in contexts where these data collection efforts take on some features of scientific research, or as was the case in the Cambridge Analytica scandal, involve entities linked to academic institutions and so gain a veneer of academic acceptability (Susser et al. 2019).

MESSAGES EMPHASISING THE BENEFITS OF PARTICIPATION: IMPACT ON CITIZEN SCIENTISTS VERSUS IMPACT ON QUIZ TAKERS

We found that the relationship between the messages presented to participants and the volume of sensitive data they disclosed was mediated by the type of project that these participants signed up for in the first place.

In a citizen science context, there were no significant differences between the Message groups, suggesting that the type of post-recruitment messages seen by the participants (learning about science, learning about self, or control) did not influence data disclosure. This is surprising as, in a previous study, citizen scientists prompted to

learn about science shared a larger volume of data when compared with citizen scientists encouraged by other messages (Rudnicka et al. 2019); this study, however, did not include a control group. It is possible that, while messages highlighting different types of potential benefits to citizen science participation can have varying impacts on citizen scientists' disclosure behaviour, they do not independently increase the volume of data disclosure. These findings support previous appraisal of data disclosure among citizen scientists in relation to Nissenbaum's contextual integrity framework (Bowser et al. 2017), suggesting that when considering data disclosure in citizen science, we must first focus on the specificity of citizen science projects and how participating in citizen science (as opposed to a different type of project involving data disclosure) may influence willingness to share information. Current results suggest that Nissenbaum's framework may be more applicable to data disclosure in citizen science than, for example, the privacy calculus framework, which focuses on rewards that motivate people to share data.

The use of messages that emphasise benefits of participation may still be needed in citizen science, as they can help support the recruitment of participants (August et al. 2018). To facilitate the design of projects that enable collection of personal and/or sensitive data, there is a need for more research that examines how recruitment messages can impact citizen scientists' willingness to share information, and how best to combine the recruitment and disclosure needs in research projects that require both the collection of sensitive data and the participation of a large sample of participants.

In the online quiz group, the participants disclosed a significantly smaller volume of sensitive data when they were presented with a message that encouraged them to discover their sleep score (learn-about-self message), than participants encouraged to learn about health psychology (learn-about-science message). While the Participant Information Sheet specified that we were studying the connections between stress and sleep (and therefore justified the presence of both types of questions, sensitive and neutral), the Neutral Item questions had more overt relevance to the study of sleep. Individuals prompted to discover a sleep score may have been keen to find it out fast and therefore paid less attention to reading the Participant Information Sheet. They could have been less accepting of sensitive data requests as a result. This is in line with previous research, which demonstrated that recruitment messages advertising self-learning tend to attract participants motivated to take part out of boredom (August et al. 2018), which, as the authors emphasise, can result in a less attentive sample (Jun, Hsieh, and Reinecke 2017).

Furthermore, research has shown that if the reward from disclosure does not match the cost of disclosure, people may choose to withhold data (Sheehan and Hoy 1999). It is possible that discovering a personal score was a less attractive reward than learning about science, causing participants in the online quiz group to be less accepting of requests for sensitive data than the participants in the citizen science group.

Projects that aim to recruit participants by drawing their attention to self-learning opportunities may benefit from emphasising the other aspect of learning, that is, acquisition of knowledge about science, post-recruitment, to facilitate the collection of sensitive data.

LIMITATIONS

While this study provides further insights into the relationship between information communicated to participants about an online survey and these participants' willingness to disclose sensitive data, these findings cannot be applied to all citizen science settings. Many citizen science projects recruit participants from existing communities, built around an interest in a topic or an interest in citizen science. Additionally, the majority of citizen scientists merely "dabble" (Eveleigh et al. 2014) in projects, before dropping out (Ponciano and Brasileiro 2014), with attrition especially prevalent in long-term projects (Rotman et al. 2012). The current study recruited citizen scientists from amongst Twitter users interested in taking a citizen science survey about sleep habits. While these individuals have some things in common (for example, interest in sleep research, or even interest in citizen science in general), they were not members of any existing community. Therefore, although this study provides insights pertinent to the participation of dabblers, which is relevant to a large number of citizen science projects, further research is needed to clarify how different populations of citizen scientists may differ in the way they approach data disclosure.

The current study generates insights based on the behaviour of participants recruited for one-off contributions, and so the findings from this research have more relevance to new and casual, rather than experienced, citizen scientists.

Finally, although the difference in the disclosure of sensitive data between the "online quiz" and "citizen science" projects was statistically significant, in practice, both groups answered only a small number of sensitive data queries and rejected many other sensitive questions. While this study highlights the role of how unpaid surveys are advertised in eliciting data disclosure, it does not answer all questions about sensitive data disclosure in citizen science. It merely highlights that whilst emphasising the brand of citizen science may increase participants' willingness to

share sensitive data, there are still areas where participants may be reluctant to share their data even with citizen science projects. There are also likely to be other factors at play, for example, whether participants see data requests as reasonable and relevant in the context of a particular research project. These factors ought to be explored in future research, ideally obtaining post-survey qualitative feedback from participants (e.g., interviews) to fully understand their rationale for making specific decisions about what data to disclose or retain. At this point, one key practical takeaway for practitioners is demonstrating the discrepancy between data disclosure and ability to retain participants. For example, researchers may wish to prioritise data disclosure or recruitment of a sizeable number of participants, and then respectively frame their project as citizen science or an online quiz. However, our study merely points to the general direction of changes in disclosure behaviour, and as the differences between groups in the current study were small, more research is needed to clarify the practical impact of the current findings on the practice of citizen science and ability of project hosts to elicit the disclosure of sensitive data when their projects require that. The impact, however small, that describing a project as citizen science has on disclosure of sensitive data does reinforce the need to ensure that users understand what the benefits of taking part in a particular citizen science project are.

CONCLUSION

This study demonstrated that when communicating to potential participants about the benefits of taking part in a project, it is not only the content of the message that influences their behaviour, but also the stage at which that message is communicated. We found that people tend to share a higher volume of sensitive data when they believe that they are taking part in a citizen science project, compared with people who believe they are filling in an online quiz focused on allowing them to discover a personal score. This demonstrates that recruitment messages can influence not only recruitment and retention of participants (as was the focus of previous research) but also the volume of sensitive data disclosed within projects.

Moreover, we found that the effectiveness of post-recruitment messages was influenced by the type of project the participants signed up for. While these messages did not have a significant impact on sensitive data disclosure among citizen scientists, quiz takers prompted to learn about science answered more Sensitive Item questions than quiz takers prompted to learn about themselves.

It appears that while emphasising self-discovery can increase the number of participants willing to take part in an online psychological survey, citizen science projects that require the disclosure of sensitive data should instead focus on emphasising the brand of citizen science at the recruitment stage.

DATA ACCESSIBILITY STATEMENTS

Data used in the research project has not been made available as consent for this purpose had not been obtained from participants.

SUPPLEMENTARY FILE

The supplementary file for this article can be found as follows:

- **Survey Questions.** Sensitive items and Neutral items used in the survey are listed here. DOI: *https://doi.org/10.5334/cstp.440.s1*

ETHICS AND CONSENT

We received ethical approval from the anonymised departmental ethics committee (UCLIC Ethics Committee Approval ID Number: UCLIC/1718/001/Staff Cox/ Rudnicka). We obtained consent from the participants by instructing them to read a Participant Information Sheet and answer one main and fifteen additional consent questions.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

All authors participated in the development of the ideas presented in this paper.

Author 1 wrote the first draft. Authors 2 and 3 contributed to the redrafting of the paper.

AUTHOR AFFILIATIONS

Anna Rudnicka *orcid.org/0000-0002-8236-6270*
University College London, GB

Sandy Gould *orcid.org/0000-0003-0476-4270*
Cardiff University, GB

Anna Cox *orcid.org/0000-0003-2231-2964*
University College London, GB

REFERENCES

Aspler, S. 2019. Customize your cheesecake and we'll reveal which high school clique you belong in, 29 June 2019. Available at: *https://www.buzzfeed.com/sarahaspler/bake-a-cheesecake-and-well-reveal-which-high-school-clique* (Last accessed 27 November 2019).

Auerbach, J, Barthelmess, EL, Cavalier, D, Cooper, CB, Fenyk, H, Haklay, M, Hulbert, JM, Kyba, CCM, Larson, LR, Lewandowski, E and **Shanley, L.** 2019. The problem with delineating narrow criteria for citizen science. *Proceedings of the National Academy of Sciences*, 116(31): 15336–15337. DOI: *https://doi.org/10.1073/pnas.1909278116*

August, T, Oliveira, N, Tan, C, Smith, N and **Reinecke, K.** 2018. Framing Effects: Choice of Slogans Used to Advertise Online Experiments Can Boost Recruitment and Lead to Sample Biases. In: *Proceedings of the ACM on Human-Computer Interaction*, New York, NY, November 2018, (CSCW): 1-19. DOI: *https://doi.org/10.1145/3274291*

Barker, D. 2014. BuzzFeed is Watching You, 24 June 2014. Available at: *https://barker.co.uk/buzzfeediswatching* (Last accessed: 14 September 2020).

Bisceglio, P. 2017. The dark side of that personality quiz you just took. *The Atlantic*, 13 July [online access at *https://www.theatlantic.com/technology/archive/2017/07/the-internet-is-one-big-personality-test/531861/* last accessed 16 December 2020].

Bonney, R, Cooper, CB, Dickinson, J, Kelling, S, Phillips, T, Rosenberg, KV and **Shirk, J.** 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, 59(11): 977–984. DOI: *https://doi.org/10.1525/bio.2009.59.11.9*

Bösch, C, Erb, B, Kargl, F, Kopp, H and **Pfattheicher, S.** 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. *Proceedings on Privacy Enhancing Technologies*, 2016(4): 237–254. DOI: *https://doi.org/10.1515/popets-2016-0038*

Bowser, A, Shilton, K, Preece, J and **Warrick, E.** 2017. Accounting for privacy in citizen science: Ethical research in a context of openness. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2124–2136. Seatthe, WA, February 2017. DOI: *https://doi.org/10.1145/2998181.2998305*

Bowser, A, Wiggins, A, Shanley, L, Preece, J and **Henderson, S.** 2014. Sharing data while protecting privacy in citizen science. *Interactions*, 21(1): 70–73. DOI: *https://doi.org/10.1145/2540032*

Domroese, MC and **Johnson, EA.** 2017. Why watch bees? Motivations of citizen science volunteers in the Great Pollinator Project. *Biological Conservation*, 208(2017): 40–47. DOI: *https://doi.org/10.1016/j.biocon.2016.08.020*

ECSA (European Citizen Science Association). 2015. Ten Principles of Citizen Science. Berlin. *http://doi.org/10.17605/OSF.IO/XPR2N*

European Parliament and Council of European Union. 2016. Regulation (EU) 2016/679. Available at: *https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN* (Last accessed: 20 December 2020).

Eveleigh, A, Jennett, C, Blandford, A, Brohan, P and **Cox, AL.** 2014. Designing for dabblers and deterring drop-outs in citizen science. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2985–2994. Toronto, Ontario, April 2014. DOI: *https://doi.org/10.1145/2556288.2557262*

Gómez-Barroso, JL. 2018. Experiments on personal information disclosure: Past and future avenues. *Telematics and Informatics*, 35(5): 1473–1490. DOI: *https://doi.org/10.1016/j.tele.2018.03.017*

Harding, X. 2018. It's not just Facebook — online quizzes also collect data about you. *Mic*, 31 March [online access at *https://www.mic.com/articles/188678/its-not-just-facebook-online-quizzes-also-collect-data-about-you* last accessed 16 December 2020].

Hern, A. 2018. Cambridge Analytica: how did it turn clicks into votes? *The Guardian*, 6 May 2018 [online access at *https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie* last accessed 20 December 2020].

Horne, JA and **Östberg, O.** 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Internacional Journal of Chronobiology*, 4: 97–110.

Jennett, C, Furniss, DJ, Iacovides, I, Wiseman, S, Gould, SJ and **Cox, AL.** 2014. Exploring citizen psych-science and the motivations of errordiary volunteers. *Human Computation*, 1(2): 201–220. DOI: *https://doi.org/10.15346/hc.v1i2.10*

Jun, E, Hsieh, G and **Reinecke, K.** 2017. Types of motivation affect study selection, attention, and dropouts in online experiments. In: *Proceedings of the ACM on Human-Computer Interaction*, 1: 15. CSCW, Article 56, November 2017, DOI: *https://doi.org/10.1145/3134691*

Lee, TK, Crowston, K, Harandi, M, Østerlund, C and **Miller, G.** 2018. Appealing to different motivations in a message to recruit citizen scientists: Results of a field experiment. *Journal*

of Science Communication*, 17(1): A1–A1. DOI: *https://doi.org/10.22323/2.17010202*

Malheiros, M, Brostoff, S, Jennett, C and **Sasse, MA.** 2013. Would you sell your mother's data? Personal data disclosure in a simulated credit card application. In: *The Economics of Information Security and Privacy*. Berlin, Heidelberg: Springer. 237–261. DOI: *https://doi.org/10.1007/978-3-642-39498-0_11*

Malheiros, M, Preibusch, S and **Sasse, MA.** 2013. Fairly truthful: The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In: Huth, M, Asokan, N, apkun, S, Flechais, I and Coles-Kemp, L. (eds.), *Trust and Trustworthy Computing. Trust 2013. Lecture Notes in Computer Science*, 7904: 250–266. Springer, Berlin, Heidelberg. DOI: *https://doi.org/10.1007/978-3-642-38908-5_19*

Oberhauser, K and **LeBuhn, G.** 2012. Insects and plants: Engaging undergraduates in authentic research through citizen science. *Frontiers in Ecology and the Environment*, 10(6): 318–320. DOI: *https://doi.org/10.1890/110274*

Ofcom (The Office of Communications). 2017. *Adults' Media Use and Attitudes Report 2017*. London, England: Ofcom.

Perkel, JM. 2020. Mischief-making bots attacked my scientific survey. *Nature*, 579: 461. DOI: *https://doi.org/10.1038/d41586-020-00768-0*

Ponciano, L and **Brasileiro, F.** 2014. Finding volunteers' engagement profiles in human computation for citizen science projects. *Human Computation*, 1(2): 245–264. DOI: *https://doi.org/10.15346/hc.v1i2.12*

Rotman, D, Preece, J, Hammock, J, Procita, K, Hansen, D, Parr, C, Lewis, D and **Jacobs, D.** 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the 15th ACM conference on Computer supported cooperative work & social computing*, 217–226. Vancouver, Canada, on March 15–18, 2012. DOI: *https://doi.org/10.1145/2145204.2145238*

Rudnicka, A, Cox, AL and **Gould, SJJ.** 2019. Why Do You Need This? Selective Disclosure of Data Among Citizen Scientists. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems – CHI '19,* 1–11. New York, NY on 4–9 May 2019. DOI: *https://doi.org/10.1145/3290605.3300622*

Sheehan, KB and **Hoy, MG.** 1999. Flaming, complaining, abstaining: How online users respond to privacy concerns. *Journal of Advertising*, 28(3): 37–51. DOI: *https://doi.org/10.1080/00913367.1999.10673588*

Silvertown, J. 2009. A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9): 467–471. DOI: *https://doi.org/10.1016/j.tree.2009.03.017*

Susser, D, Roessler, B and **Nissenbaum, H.** 2019. Technology, autonomy, and manipulation. *Internet Policy Review*, 8(2): 1–22. DOI: *https://doi.org/10.14763/2019.2.1410*

Theobald, EJ, Ettinger, AK, Burgess, HK, DeBey, LB, Schmidt, NR, Froehlich, HE, Wagner, C, HilleRisLambers, J, Tewksbury, J, Harsch, MA and **Parrish, JK.** 2015. Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research. *Biological Conservation*, 181: 236–244. DOI: *https://doi.org/10.1016/j.biocon.2014.10.021*

Thomas, SM and **Jha, R.** 2017. How Privileged Are You, 21 August 2017. Available at: *https://www.buzzfeed.com/soniathomas/how-privileged-are-you* (Last accessed: 27 November 2019).

UK Government. 2018. Data Protection Act 2018. Available at: *https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted* (Last accessed 20 December 2020)

Wiseman, S, Cox, AL, Gould, SJJ and **Brumby, DP.** 2017. Exploring the effects of non-monetary reimbursement for participants in HCI research. *Human Computation*, 4(2). DOI: *https://doi.org/10.15346/hc.v4i1.75*

Zuckerberg, M. 2018. 'Facebook Post'. Available at: *https://www.facebook.com/zuck/posts/10104712037900071* (Last accessed: 20 December 2020).